Fast Frame Hybrid Optoelectronic Neural Object Recognition System

Tien-Hsin Chao, Taher Daud, and Anilkumar Thakoor
Center for Space Microelectronics Technology
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099

Key Words: Target recognition, optoelectronic, neural networks, hardware implementation

Abstract

The Hybrid Optoelectronic Neural Object Recognition System (HONORS) development at the Jet Propulsion Laboratory (JPL) promises high speed (>1000 frames/s) processing of information obtained with large frame size (1000x1000-pixel) images. It consists of two major building blocks: (1) an advanced grayscale optical correlator (OC); and (2) a massively parallel, VLSI-based neural 3-D processor (N3DP). The OC, with its inherent parallel processing and shift invariance, is used for target of interest detection and segmentation. The N3DP, with its high speed convolution operations (10^{12} operations/s) and neural learning capability, is used for target classification and identification. This paper presents the system architecture and processing algorithms. In addition, the results from simulations and experiments, including the detection, classification, and tracking of tanks and aircraft are summarized.

Introduction

Automatic target recognition (ATR) applications require high speed processing, recognition of objects from cluttered background, and intelligent decision making. A compact, high speed ATR system has always been an illusive, moving target. Enhancement in processor speeds have gone hand in hand with better and larger sensors and imagers requiring even higher data processing rates and autonomy in processing and decision making. Various civil, military, and space applications would see orders of magnitude advancement in their capabilities with a system performing ATR at 1000 frames per second especially with large format (1000x1000-pixel) imagers.

JPL is developing a hybrid optoelectronic neural object recognition system (HONORS), for high-speed detection, segmentation, classification and identification of objects from noisy/cluttered backgrounds. Unique advantages of HONORS include high-speed (<1 ms per frame), large input frame (1000 x 1000-pixel), high discrimination accuracy (> 90%), and ease of training. HONORS consists of an optical correlator (OC) and a neural 3-dimensional processor (N3DP). The OC consists mainly of a unique gray-scale spatial light modulator (SLM) as the high resolution correlation filter and is used for object detection and segmentation. Due to the inherent parallel processing capability of OC, it would perform wide area surveillance for target of interest (TOI) in less than 1 ms per frame. The detected and segmented TOI would be handed over to the N3DP.

The N3DP consists of a 64-channel high-speed electronic convolver coupled to a multilayer electronic neural network. Each input (64x64-pixel) would be simultaneously mapped to 64 eigenvector-based object data bank images. The output from each input image would be a 64-element feature vector. The electronic neural network would subsequently classify the input feature vector amongst multiple classes of objects.

Both the correlation filter and the eigenimage data bank rely on training from example images of known classes. Training relies on rules developed using optimization process. More specifically, a Maximum Average Correlation Height (MACH) algorithm is used for correlation filter training. Eigenimage computation is used to establish object data bank.

System Architecture

The system block diagram of HONORS is shown in Figure 1. The entire ATR function is implemented in 5 consecutive steps [1]:

(1) Sensor data/images, acquired by a multisensor platform in both 1-D (hyperspectral, acoustic) and 2-D (focal plane arrays, synthetic aperture radar, etc.) formats are fed into HONORS through a frame buffer (FB) device.

(2) After format conversion, FB will feed the full frame input into the succeeding OC for preprocessing for target of interest (TOI) detection and segmentation based on

target shape, size, texture, and gray-scale distribution. Training target images are computed with a distortion invariant correlation filter algorithm and downloaded into the optical correlator. It is also highly resistant to background noise and clutter. Due to its inherent parallel processing and shift invariance capabilities, the OC would be used for wide-area survey. The output of the OC preprocessor will be a list of targets (including both true and false targets) marked with their locations and types (e.g. tank, truck, missile, etc.).

(3) Based on the TOI data output from the OC, a column loading input chip (CLIC) will acquire the segmented TOI images saved in FB.

(4) In addition to CLIC, N3DP has two more building blocks: an eigen-vector-based feature extraction 3-D stacked electronic processor and an analog neural network classification chip. The N3DP, properly trained with a target database, will perform final target classification and identification.

(5) The output of N3DP will be a viable input for target tracking, navigation & guidance, sensor retasking, and mission replanning. Details of the two building blocks of HONORS: OC and N3DP are described in the following sections.

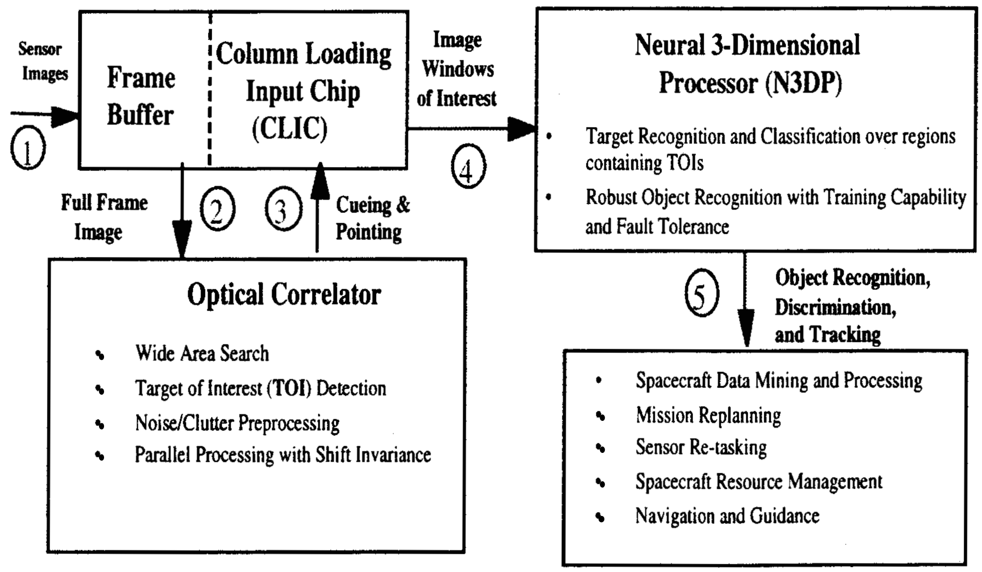


Figure 1. System Block Diagram of HONORS for ATR

JPL has recently developed a high-speed camcorder-sized grayscale optical correlator [2-4] consisting of a grayscale input SLM to replace the binary SLM used in previous state-of-the-art optical correlator systems. A ferroelectric liquid crystal (FLC) SLM capable of encoding real-valued data is also used for the correlation filter implementation.

A system architecture for this innovative optical processor is shown in Figure 2(a). A 25 mw diode laser, emitting at 690 nm, is collimated and used as the light source. A 640 x 480-pixel liquid crystal display (LCD) is used as the input SLM. A 128x128-pixel FLC SLM is used as the filter SLM. The LCD SLM operates in a transmission mode while the FLC SLM operates in a reflection mode. The input image is Fourier-transformed and directed to address the filter SLM via a polarizing cubic beam splitter. A half-wave plate is inserted between the beam splitter and the filter SLM to steer the polarization orientation of the throughput light beam to ensure bipolar-amplitude output from the FLC SLM. The readout beam, reflected back from the filter SLM is inverse Fourier-transformed at the output correlation plane. A high-speed CCD is used to grab the output for peak-detection. A photograph of this grayscale optical correlator is shown in Figure 2b.

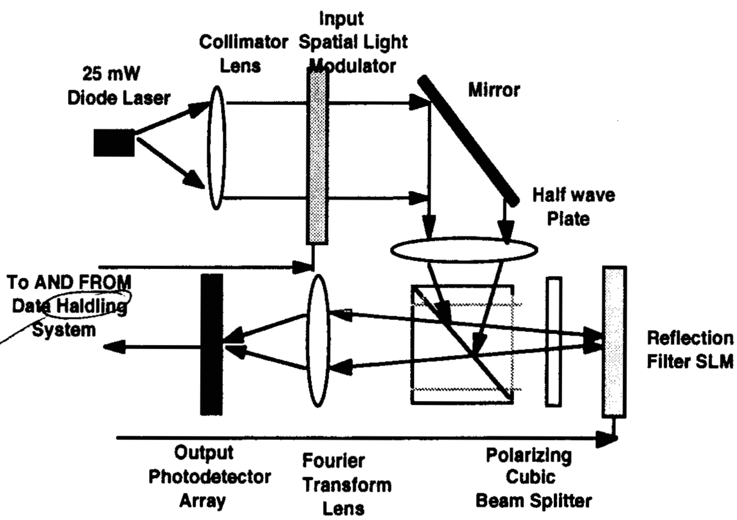


Figure 2a

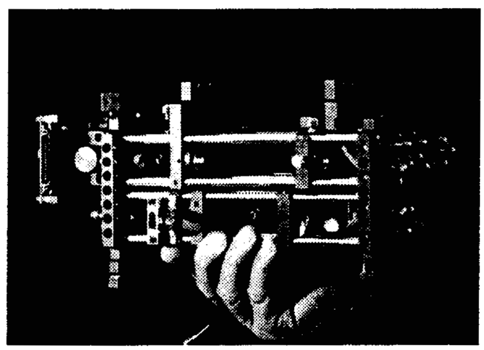


Figure 2b

Figure 2. Compact grayscale optical correlator: (a) Schematic diagram; and (b) Photograph of the palmcorder-size correlator system.

The input LCD SLM possesses 640x480 pixels with a 24 mm pitch. This results in a 15.4 mm x 11.2 mm aperture, very suitable for compact packaging. The contrast ratio is about 100:1 with an 8-bit grayscale resolution when used for imaging display.

The SLM is addressed in a VGA graphic mode. The key device leading to the real-valued filter modulation is the 128x128 analog FLC SLM built using liquid crystal on silicon technology [1]. It utilizes high-tilt FLC material resulting in the use of all positive real amplitudes, binary phase-only, and bipolar-amplitude modulations easily obtained by varying the orientation of the halfwave plate placed in front of the SLM. The FLC SLM has a switching time of 50 to 100 ms that provides a frame rate up to 10 kHz. The contrast ratio is 76:1. An estimated 4-6 bit output resolution can be achieved by using an 8-bit gray-scale OC resolution.

Optical Correlator Filter Algorithm

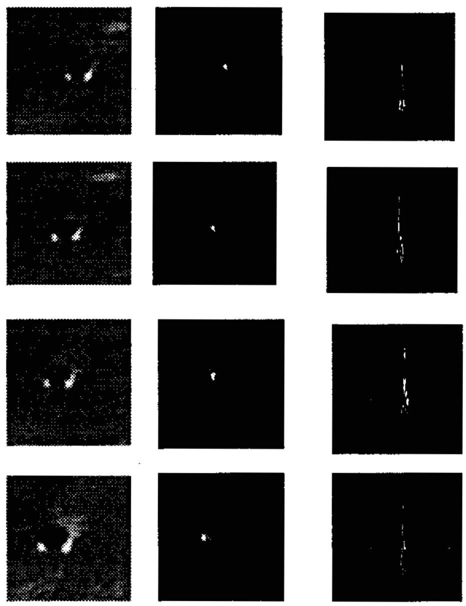
The unique real-valued filter modulation capability of the grayscale OC has enabled us to select a more robust correlator filter algorithm for optical implementation. A distortion invariant correlator filter algorithm, MACH (*maximum average correlation height*) [1,5-7], has been selected and implemented. An overview of the MACH filter algorithm has been given elsewhere [1,7].

Basically, the filter is synthesized as a composite of several training images (samples picked from the target database). A MACH filter is desired to generate an output correlation plan that has a high-peak response with respect to all input images from the same class. To compute the MACH filter transfer function, h , a mean square error criterion referred to as the *average similarity measure* (ASM) is used as a metric for distortion. The filter response is more invariant with respect to a smaller ASM. Therefore, in filter design, it is required that h have a high correlation peak with the mean image while making the ASM small. It is also required to possess noise tolerance to reduce the output noise variance (ONV).

The filter h maximizes the height of the mean correlation peak relative to the expected distortion. The superior performance of MACH filter is attributed to the inclusion of the ASM criterion which reduces the filters' sensitivity to distortions and to the removal of hard constraints on the peak. This permits the correlation planes to adjust to whatever value best permits the optimization of performance criterion.

Optical Mach Filter Demonstration Of Distortion-Invariant Target Detection

In a laboratory experiment for distortion invariant target detection using optically implemented MACH filter, a sequence of 15 IR images of a tank moving downward a desert terrain is used as the input. In preparing the MACH filter, 5 of the inputs were selected for training and the rest used for testing. A MACH filter is computed and then downloaded into the filter SLM. In Figure 3, results of the optical correlator output are provided. All 15 inputs were successfully detected with one single MACH filter trained for recognizing the tank images. Four out of the 15 total input images used are shown in Fig. 3 (a), and their corresponding correlation peaks and 3-D plots are shown in (b) and (c) respectively. Notice that the correlation peaks remain sharp and uniform across all the input images including those with scale and perspective variations. The scale ratio between the top and bottom images is about 2:1.



(a) (b) (c)

Figure 3. Target detection results using an optical MACH correlator filter.

This experiment validates that an optical correlator is an effective target detection preprocessor capable of accomplishing target detection and segmentation. Using a robust MACH filter will enable distortion invariant, noise resistant filter training to drastically reduce the number of false alarms sent into the subsequent classification/identification stage.

Neural 3-Dimensional Processor (N3DP)

N3DP is a ground and airborne image-processing/target-recognition processor being developed as an enhanced version of a 3-dimensional artificial neural network (3DANN) processor architecture.

ARCHITECTURE: A block diagram of the front part of the N3DP consisting of a CLIC and an NPM cube (without the subsequent multilayer neural network) is shown in Figure 4. It also shows a photograph of the hardware. The function of N3DP is to have a 64x64-pixel image window as an electronic input to be

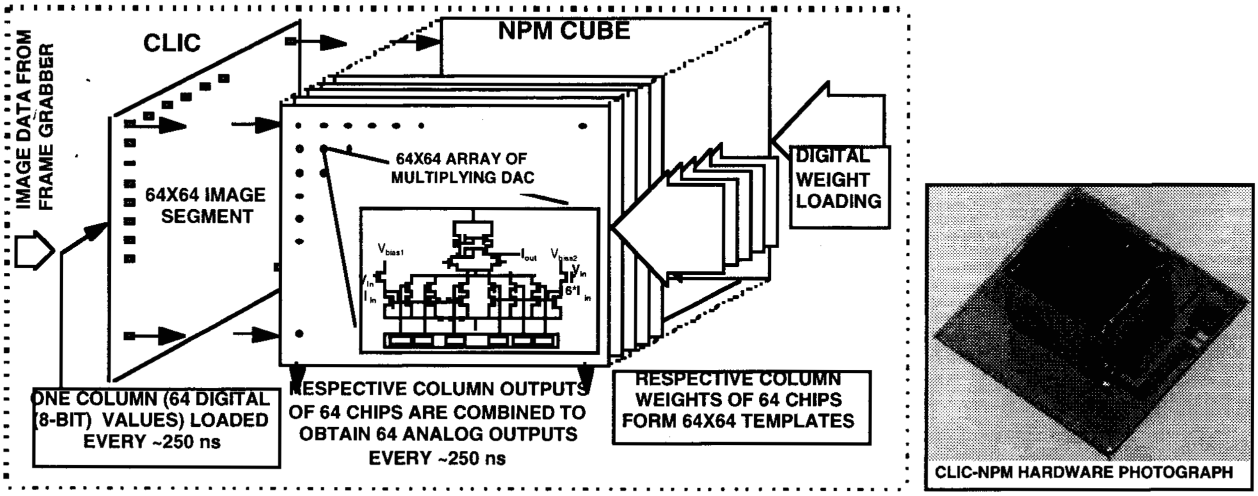


Figure 4. The 3D convolver network consists of a neural processing module (NPM cube) with 3-D stacked 64 chips, each with a 64x64 synapse array based on 8-bit multiplying digital to analog technology and incorporating a special-purpose image-write device called CLIC that is bump-bonded to the NPM cube. It has been realized (photograph) as a 10-gm, 3-cm^3 package, with power consumption of 2.5 W.

convolved simultaneously with 64 selected and stored templates (each of 64x64-pixel size) and provide the 64 electronic outputs (one innerproduct per template) at a speed of about 4 MHz, reducing the input bandwidth of 4096 to just 64 outputs. N3DP consists of three components: (I) a CLIC; (ii) a set of 64 neural processing (with an architecture of a 64x64 array of multiply-accumulate circuits) chips stacked as 3D cube (with edge-inputs) termed as neural processing module (NPM) mounted on a mother board; and (iii) a multilayer neural network. A 64x64-pixel image window as digital input can be fed to the CLIC, one column at a time.

CLIC consists of a set of 8-bit shift registers and a 64x64 array of 8-bit static random access memory (SRAM) cells attached to respective multiplying digital to analog converters (MDAC). Further details of the CLIC and the chip test results are provided elsewhere [8]. The inputs are thus converted by MDACs to 4096 analog voltage signals, and fed in parallel into the NPM cube every 250 nanoseconds.

Within each of the 3D stacked NPM chip circuits, the chief component of each cell of the 64x64 array is a high-speed multiply-accumulate (synapse) circuit [9]. The circuitry is designed to operate at a low power consumption of only 3 to 5 μwatts, or <1 watt per cube. The templates are provided by the digitally stored 8-bit weights in the synapse SRAM circuit. Each template is stored column-wise on all 64 chips, first column of the 64 chips providing the first stored template, etc. The 64 sets of neural computations with 4096 (i.e., 64x64) analog inputs can be accomplished in 250 nanoseconds [10] (i.e., 10^{12} multiply and add operations in one second).

<u>NEURAL ALGORITHMS, SIMULATION, AND ANALYSIS:</u> Work on neural algorithm development, simulation, and analysis has been done using this architecture. The NPM network produces 64 inner-products (one 4096-element input with 64, 4096-element template vectors). Originally with the concept of rastering the large image, it is to be fed by inputting a new column or row of a 64x64 subwindow to CLIC every 250 ns (at start, 63 column-input operations would fill the CLIC register array) accomplishing 64 convolutions of a 256x256 image with 64x64 masks in just 16 ms.

The 64 analog values generated by NPM every 250 ns are converted to 8-bit digital values and passed along to the associated memory and the point operation processor (POP) programmed as a neural network. A custom VLSI implementation of POP could be designed and fabricated; however, presently a commercial parallel processing board is being used. POP performs the desired target recognition/tracking functions on the NPM output. Command and control of various operations (e.g., detection/ classification/tracking mode command, loading of templates, point operation functions, data recording, etc.) are done through a host machine.

To achieve target recognition of objects of arbitrary size and orientation efficiently, a hierarchical neural network approach based on eigenvectors is employed, as shown in Figure 5. Using NPM as the dedicated synapse weight multiplier hardware, 64 eigenvector templates representing the principle axes of a collection of multidimensional data points (i.e., object images of various configurations) have been employed [11-13]. Since each data point (image) is a 4096-element vector, finding a set of 4096 orthonormal eigenvectors is possible (64 of which can reside on NPM cube at one time). Selecting the top 64 eigenvectors derived from principal component analysis of target imagery reduces the image dimensionality while retaining much of the relevant classification information.

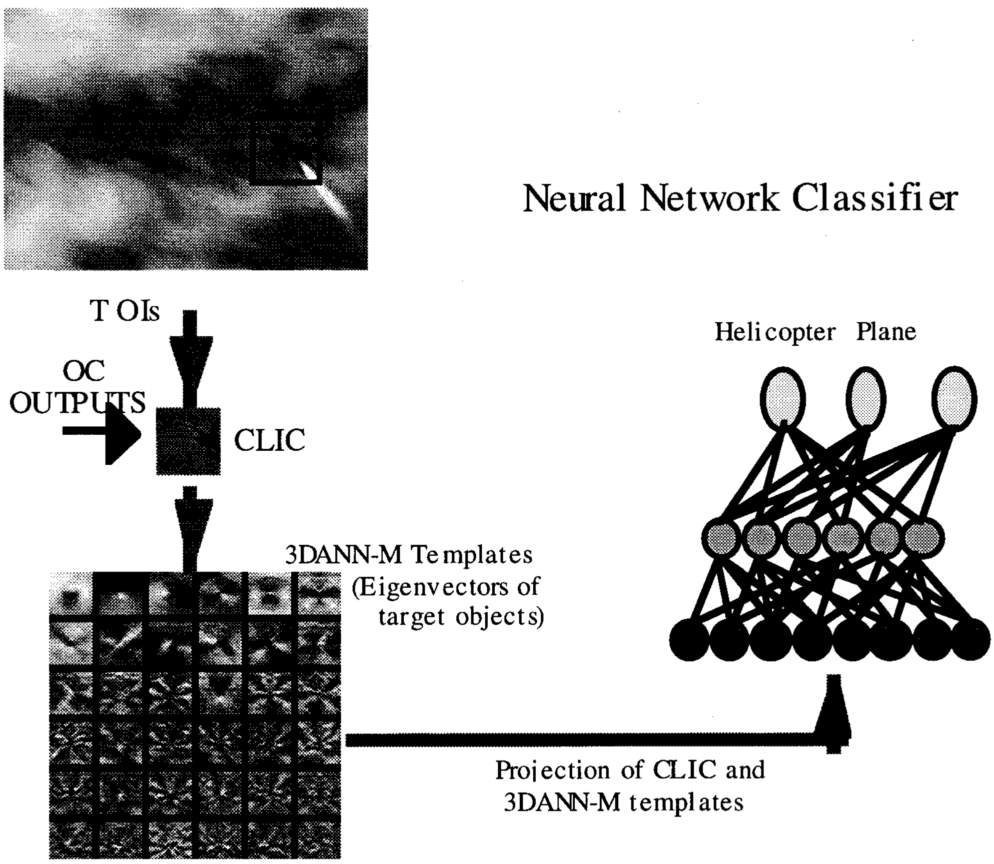


Figure 5. General target recognition is achieved using eigenvector projections in conjunction with a neural network classifier trained on selected data sets.

The problem that has plagued pattern recognition systems is that, unless some restrictions are placed on variations in the target imagery, the top components become quite general and perhaps unsuitable for fine distinction of a target with all possible scale, rotation, and perspective variations. Our strategy is to parameterize (e.g., lighting, pose, class, identity, and scale) and partition the object space in a hierarchical fashion. The top layer is the detection layer, trained to represent the presence or the absence of a target. Then, each partition is first used to detect and segment all possible targets (with some false alarms). To classify each partition, a neural network is trained on data imagery drawn from the set of variables that define the partition and projected onto eigenvectors suitable for that particular distribution of data.

Information about the object (its class, identity, or orientation, etc.) is processed in a coarse-to-fine manner. For instance, after detecting an object in a frame, a rough estimate of image orientation/scale is made, a result that can then be used to limit the variation that needs to be considered during subsequent object classification step. In simulation, object recognition rates in cluttered background of 96% have been obtained [13].

Conclusions

HONORS, a powerful ATR system described herein combines an advanced optical correlator and a 3-D integrated neural network based processor in a compact package to perform object recognition with unprecedented speed. Algorithms have been identified and simulated for the optimum operation of both the optical correlator and N3DP. Demonstrations for real-time detection, classification, and precision-tracking, with ground-based and airborne experiments on live targets are planned. It can be projected that such a high performance system will find varied uses both in the NASA arena and for commercial and military applications.

Acknowledgments

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was jointly sponsored by the Ballistic Missile Defense Organization and the National Aeronautics and Space Administration. The authors wish to express their sincere thanks to S. Udomkesmalee, T. Duong, C. Padgett, and G. Reyes for helpful technical discussions.

References

1. T-H. Chao, T. Daud, and A. Thakoor, "Honors: Advanced Miniature Processing Hardware for ATR Applications," Proceedings of the 4th International NEURAP Conf., March 11-13, 1998, Marseille, France. (In Press)
2. T-H. Chao, E. Hegblom, B. Lau, and W. J. Miceli, "Optoelectronically Implemented Neural Network with a Wavelet Preprocessor," *Proc. of SPIE* vol. 2026, 1993, San Diego, CA., pp. 472-482.
3. T-H. Chao and A. Yacoubian, "Real-time updatable optical wavelet processor," *Proc. of SPIE*, vol. 2490, 1995, pp. 118-124,.
4. T-H. Chao, A. Yacoubian, and B. Lau, "Portable optical wavelet processor," *Proc. of SPIE*, vol. 2752, 1996, pp. 254-259,.
5. B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.*, vol. 31, no. 23, pp. 4773-4801, 1992.
6. A. Mahalanobis, B. V. K. Vijaya Kumar, S. Song, S.R.F. Sims and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt*, vol. 33, pp. 3751-3759, 1994.
7. A. Mahalanobis, "Correlation filters for object tracking target re-acquisition and smart aimpoint selection," *Proc. of SPIE*, vol. 3073, 1997, pp. 25-32.
8. T. Duong, T. Thomas, T. Daud, A. Thakoor, and B. Lee, "64x64 Analog input array for 3-dimensional neural network processor," *Proceedings of the 3rd International Conference on Neural Networks and Their Applications*, Marseilles, France, March 1997.
9. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, 1943
10. T. Duong, S. Kemeny, M. Tran, T. Daud, and A. Thakoor, "Low power analog neurosynapse chips for a 3-D sugarcube neuroprocessor," *Proceedings of the IEEE International Conference on Neural Networks*, vol. III, 1994, Orlando, FL, pp. 1907-1911.
11. M. Turk and A. Pentland, "Eigenfaces for recognition," *J. of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
12. C. Padgett, G. Cottrell, and R. Adolphs, "Categorical perception in facial emotion classification," *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Hilldale, 1996, pp. 201-207,.
13. C. Padgett, M. Zhu, and S. Suddarth, "Detection and object identification using VIGILANTE processing system", Proc. SPIE, vol. 3077, 1997.